면합수를 적용한 얼굴인식

김희승, 김종호
서울시립대학교 컴퓨터통계학과
hskim@uos.ac.kr

Face Recognition using Facet Function Model

Hee Sung Kim, Jong Ho Kim
Dept. of Computer Science and Statistics, University of Seoul

요 약

This paper represents a method for face recognition using a facet function. The region of face is extracted cutting off the hair part and normalized with respect to the horizontal eye line. One of the intractable problem in face recognition is to overcome the gradual change of the gray level intensity due to the direction of light. We applied facet function model on the array of the gray levels of the local area of the face image so that the facet function can reveal the slope of the local area of face. This supplies a few parameters on each local area of face which can be used for the feature vectors of the neural nets. This shows improved rate of face recognition comparing to the existing methods.

Keywords: 면합수, 역전파 다층 퍼셉트론, 주성분 분석법

1. 서론

최근 시스템 성능의 향상과 더불어 생체기관을 인증에 이용하는 생체인식이 중요한 과제로 떠오르고 있다. 생체인식은 지문인식, 화자인식, 정력인식, 얼굴인식 등이 있으며, 얼굴인식에 대한 많은 연구가 이루어지고 있다. 하지만 얼굴인식은 외부환경의 변화, 얼굴의 표정변화, 해어스타일의 변화 등과 같은 여러 가지 요인들에 의해서 실제적인 용용에 많은 어려움이 있다.

일반적으로 얼굴인식은 얼굴 추적, 얼굴 영역 추출, 얼굴 인식 기술로 분류하며, 그 중 얼굴인식 기술은 전처리, 특징추출, 인식의 세 단계로 구분할 수 있다. 전처리 과정에서는 얼굴영역의 잡음을 제거하거나 조명의 변화를 보정하고, 특징추출 과정에서는 다양한 방법에 적용할 특징을 얼굴영역에서 추출해 내며, 얼굴인식 과정에서는 추출한 특징들을 이용하여 기존의 얼굴영역과 비교를 수행한다. 기존의 연구에서는 조명의 변화를 보정하기 위해 히스토그램 평활화의 전처리 기법을 사용한 후에 PCA(Principal Component Analysis) 기법을 사용하여 얼굴을 인식한다[1].

본 논문에서는 눈의 위치를 찾아내어 그 위치를 기준으로 얼굴 영역을 잘라내어其中有의, 이 방법은 얼굴이 아닌 부분을 제거하고, 얼굴영역 확득 시 상황에 따라 얼굴영역이 위쪽 혹은 아래쪽으로 치우치는 현상은 보정할 수 있다. 이렇게 추출된 얼굴의 임정 영역에 대해 조명 변화에 견고하고, 점차적인 인식률, 수행 시간도 향상시킬 수 있는 면합수를 적용하여 특징을 추출하는 방법을 제시한다. 또한, 제안된 기법에 다층 퍼셉트론 신경망을 적용하여, 성능향상에 미치는 결과를 알아본다.

2. 관련연구

인간의 시각 시스템은 조명의 영향을 받아 그림 자를 이용하여 모양을 유추해 내는 shape from shading과정을 거치는 것으로 알려졌는데[2], 이렇게 특정 물체를 인지하기 위해서는 그 물체의 형태를 인지 유추해 내야 한다. 얼굴 인식을 하기 위해서는 먼저 얼굴 영역이 위치한 곳으로부터 얼굴 영역을 추출해 내야 한다. 이때 얼마만큼 정확히 얼굴영역
을 추출해 내는다는 이후 얼굴인식의 성능에 중요한 영향을 준다.

얼굴 영역 추출 기술은 얼굴 인식 및 표정 인식을 위한 필수적인 전처리 기술이지만, 아직까지도 많은 분야에서 적용받는 기술로써 많은 연구가 진행중인[3]

얼굴 영역 추출 방법은 피부색을 이용한 방법[4], 특정 기반 방법[5], 차원성을 이용한 방법, 특징과 피부색을 결합한 방법[6][7]등이 있다.

피부색에 기반한 방법은 단순한 영상의 경우에 높은 성공률을 보이고 있지만, 피부색과 유사한 배경이 있는 경우 추출 성공률이 현저히 떨어지는 단점을 있어 단독으로 잘 사용되지 않고 다른 방법들과 결합하여 사용된다. Ming-Hsuan Yang은 다양한 크기의 얼굴모양과 피부색을 비교하는 방법을 이용하여 크기와 방향이 다양한 얼굴영상을 검출했다[4].

특정 기반 방법은 눈,코,입 등 상관관계를 고려하여 얼굴 영역을 추출하는 방법으로 높은 성공률을 보이고 있으나 알고리즘이 복잡하고 처리시간이 많이 걸리는 단점이 있다. 장정식[5]은 혈액영양에서 눈,코,입을 추출하였는데, 먼저 라플라시안 피라미드를 구성하여 저해성도 라플라시안 영상을 구한 후, 일반정보를 이용하여 눈과 코 후보를 각각 추출하고 눈과 코가 될 수 있는 조합을 구하여 원 영상에서 이들 연관된 얼굴 구성을하는 후보 영역에서 에지(edge)를 추출한다. 각 구성을후보에 여러 후보 점들 중에서 가장 조건을 만족시키는 후보들을 선택하는 방식으로 추출할 수 있었다.

차영성지지를 이용한 방법은 프레임간의 차이 영상을 이용하여 얼굴 영역을 검출하는 방법으로 다른 방법에 비해 탐색 공간을 상당히 줄일 수 있어 처리 시간을 줄일 수 있다. 하지만 음직임 정보만으로는 충분한 정보를 얻을 수 있기 때문에 독자적으로 사용하기보다는 다른 방법들과 결합하여 사용된다. 특징과 피부색을 결합한 방법은 비교적 안정적인 추출 성능을 보이고 있는데, 유통용과 오일식[6]은 얼굴 세부 정보를 사용하여 얼굴 영역을 찾고, 얼굴 영역 내에서 눈과 입 특성을 탐지하는 방법을 이용하여 95.8%의 추출 성능을 보이고 있다.

얼굴인식은 다양한 요인들로 인해서 성공적으로 인식하는데 어려움이 있다. 시각된 사람을 매우 흔한 얼굴을 하고 있는 경우도 있고, 같은 사람이라도 빛이나 영동 표정, 안경 착용, 헤어스타일등에 의해서 아주 다르게 보이기도 하기 때문이다.

최근의 얼굴인식 기법은 크게 기하학적인 특징 정합 방법과 패턴매트 정합 방법으로 나눌 수 있다. 기하학적 특징 정합 방법은 눈,코 등 특징 백터를 추출하여 두 얼굴의 유사도를 비교하는 방법으로 얼굴 인식 연구에 많은 시간이 소요되며, 각각의 얼굴에서 특징 백터를 추출하는 데 많은 어려움이 있다. 덴플롯 패턴 정합 방법으로는 주성분 분석법 (PCA)[8], 국부적 특징 분석 (LFA Local Feature Analysis), 선 성분에 기반한 얼굴인식[9], 신경망을 이용한 얼굴인식 방법이 연구되고 있다. 주성분 분석법 (PCA)은 백터 표현의 통계학적 특성을 이용하는 방법으로, 전체 영상의 데이터로 이루진 공분산 행렬에서 고유벡터와 고유치를 구하고, 고유벡터를 2차원 영상으로 변환하면 고유 얼굴(eigenface)을 구할 수 있는데, 이렇게 구한 고유얼굴들과 입력된 실시간영상과의 유사도를 측정하여 얼굴 인식을 수행한다[8].

그림 1. PCA에서 사용하는 고유얼굴

그리나 PCA 기법은 얼굴 표정, 각도 등에 안정적인 성능을 보이지만, 절연에 대한 많은 제한을 요구하고 시간비용이 높고, 모델변화에 대한 견고하지 못한 단점이 있다.

Stefan Aebisher and Olivier de Vél[9]는 선 성분에 기반한 얼굴인식 방법을 제안했는데, 얼굴영역에 많은 수의 임의의 선을 고르고, 선을 일정한 치수로 가진 격자로 표현(lattice line)하며, 각 격자 내부 색별들의 영등도를 소요로 하는 백터로 선을 표현한다. 입력된 실험 영상에서 임의의 수의 선들과 데이터베이스에 저장된 선들 중 가장 유사한 선을 선택해서 신뢰도를 구하고, 가장 높은 신뢰도는 높은 클러스터를 실험 영상의 클러스터로 선택하는 방법이다. 이 방법은 하나의 실험 영상이 아니라 여러 개의 실험 영상에 대한 클러스터를 구할 때 높은 성공률을 보이고 있다.


형광창점은 구현이 쉽고 비교적 좋은 성능을 나타내지만, 얼굴의 기울이기이나 표정변화 등에 민감하
다. 이에 비해 신경망은 적절한 신경망의 구조와 파라미터들을 구하기 쉬우지만, 영상의 변화에 대해 유연성을 가지며, 학습 후 가중치만 구성되면, 각각의 원형얼굴과 비교할 필요가 없으므로, 인식시간을 단축시킬 수 있다.

3. Haralick의 변환수 모델을 이용한 얼굴인식

조명 변화에 독립적인 얼굴인식을 위해서는 얼굴영상의 3차원적 근본구조를 표현할 수 있는 특징치의 추출이 필요하다. 영향받는 영상평량에서의 명암도 수치로부터 3차원적인 근본의 구조를 나타낼 수 있기 때문에 변환수 모델을 기반으로 특징을 추출하여 얼굴인식의 입으로 사용한다.

얼굴이 아닌 부분을 포함한 영상 전체를 입으로 사용하려면 얼굴인식의 성능에 부정적 영향을 주기 때문에 본 논문에서는 얼굴영상에서 눈의 위치를 감출하고 그 위치를 기준으로 임정 영역을 장려하는 방법을 이용함으로써 얼굴이 아닌 부분을 제외시킨 영상으로 실행을 수행하여 실험의 객관성을 높였다. 또한 제한된 방법에 수행속도를 단축시킬 수 있도록 신경망을 적응하여 본다.

3.1 Haralick의 변환수 모델

피셀 공간의 명암도로 이루어지는 고르지 않은 면에 구배(gradient)를 얻어내기 위한 변환수는 여러 가지가 있다[11], 그중 Haralick의 변환수는 영상의 구배를 얻어내어 원단선을 구하기 위해서 사용되었던 방법이다. 피셀 공간의 명암도에 적합(fit)시킬 변환수를 다음과 같이 2변수 3차 함수로 선정한다[12].

\[
h(r, c) = k_1 + k_2 r + k_3 c + k_4 r^2 + k_5 r c + k_6 c^2 + k_7 r^3 + k_8 c^3 + k_9 r^2 c + k_{10} c^3 - f(r, c)^2
\]

이 오류를 극소화시키는 \(k_1, k_2, \ldots, k_{10}\)을 구하기 위하여 오류 단순화의 힘 \(e^2\)을 \(k_1, k_2, k_{10}\)에 관하여 미분을

\[
\frac{\partial e^2}{\partial k_i} = 0
\]

한 후, \(\frac{\partial e^2}{\partial k_i} = 0\)으로 설정하고, 연립방정식을 풀면 각 계수들을 구할 수 있다.

적합한 변환수 \(h(r, c)\)의 구배(gradient)는

\[
G_h(r, c) = i \frac{\partial f}{\partial x} + j \frac{\partial f}{\partial y}
\]

로 구할 수 있고, 중심 \((0, 0)\) 폭설에서의 구배 방향 \(\alpha\)는

\[
\alpha = \tan^{-1}\left(\frac{\partial f}{\partial y}, \frac{\partial f}{\partial x}\right)
\]

에서 얻을 수 있다. 즉, 변환수에서

\[
\frac{\partial f}{\partial y} \big|_{(0,0)} = k_2, \quad \frac{\partial f}{\partial x} \big|_{(0,0)} = k_3
\]

이므로

\[
\alpha = \tan^{-1}\left(k_2, k_3\right)
\]

이다. 그리고

\[
\sin \alpha = k_2 / \sqrt{k_2^2 + k_3^2}
\]

\[
\cos \alpha = k_3 / \sqrt{k_2^2 + k_3^2}
\]

이다. 본 논문에서는 변환수에서 얻어진 구배의 방향성과 크기를 동시에 포함하는 \(\sin \alpha\) 값을 특정치로 하여 얼굴인식을 위한 입력으로 한다.

(1) 각도에 대한 차이를 구하지 않아도 되므로, 방향성에 대한 비교가 용이하다.

(2) \(\sin \alpha\) 값은 \(-1 \leq \sin \alpha \leq 1\) 사이의 값을 가지므로, 신경망의 입력(bipolar)으로 사용할 때, 다시 조정할 필요가 없다.

그림 2는 얼굴영상에서 원도우별로 변환수에서의 구배의 \(\sin \alpha\) 값을 얻어내는 과정이다.
Face Recognition using Facet Function Model

그림 2. 얼굴영상에서 변향수를 이용한 특정추출과정

3.2 얼굴 고유 영역 추출

얼굴 영상을 인식 시스템에 사용하기 위해서는 실험 영상을 일정한 형태로 정규화 시켜야 한다. 본 논문에서는 AR 데이터베이스의 얼굴 영상을 이용하여 눈의 위치를 검출하고, 그 위치를 중심으로 이바셔까지 일정한 크기로 잘라내어 실험에 사용하였다.

본 실험에서 사용하는 영상은 조명변화가 있고 표정변화가 있으며, 영상에 따라서 해어스타일이 자주 들고, 안경을 쓰고 있는 영상도 다수 포함하고 있다.

그림 3. 실험에 사용한 다양한 얼굴영상의 예

이 영상들로부터 눈의 위치를 검출하기 위해서 먼저 머리카락 부분을 포함하는 1차 얼굴 영역을 추출하고, 이치화 연산을 이용하여 머리카락을 제외한 2차 얼굴 영역을 추출한 후, y축 히스토그램의 최대 점을 이용하여 눈의 위치를 검출해낸다.

눈의 위치를 검출할 때 해어스타일에 의해 많은 영향을 받는데 이를 단순화하기 위해서는 머리카락 영역을 제외시킬 필요가 있다. 머리카락 부분을 제외한 얼굴 영상은 1차, 2차 얼굴 영역 추출 과정을 통해서 얻을 수 있다. 복잡한 배경의 영상이 아니므로 먼저 Laplacian 필터를 적용하여 경계선을 추출하고 경계선 정보를 이용하여 1차 얼굴 영역 (1차 MBR: Minimum Boundary Rectangle)을 구한다. 얼굴 영상의 경계선을 이용하여 MBR을 구하면 얼굴 영상을 포함하는 대략적인 영역을 얻을 수 있다. 이때 배경

이 덜 포함되도록 하기 위해서 원쪽, 오른쪽, 위쪽의 위치를 조금씩 더 안쪽으로 이동시켰다.

다음으로 머리카락 영역이 얼굴보다 어둡다는 점을 이용해서 2차 얼굴영역을 구하는데, 전역 이치화를 수행하고 환 폭선의 비율을 이용하여 2차 얼굴 영역 (2차 MBR)을 구할 수 있다. 조명변화가 있는 영상이 나 두두얼굴의 영상에서도 머리카락과 얼굴의 구분을 명확히 하기 위하여 전역 이치화 연산을 사용한다. 일반적으로 평균 값을 입체적으로 설정하는데 그럴 경우 머리카락 이외의 많은 부분이 포함되므로 입체를 더 냉게 설정하였다.

그림 4. 추출된 얼굴 영역

다음은 2차 얼굴 영역의 원쪽, 오른쪽 지점을 구하기 위한 조건식이다. 원쪽 지점을 구할 때는 가장 왼쪽부터 중앙 쪽으로 순차적으로 조건식을 적용하게 되는 조건이 맞는 지점을 원쪽, 오른쪽으로 선정한다. 2차 얼굴 영역은 1차 얼굴 영역을 이치화 하여, 검은 폽xED신 폹xED일의 비교한 그 영역을 구할 수 있는데, 영상의 머리카락 영역이 검은색으로 바뀌지 않고 흰색으로 바뀌어 2차 얼굴영역은 잘못 설정하게 된다. 하지만 금발이라 하더라도 얼굴과 머리카락 영역의 경계 부분은 이치화 영상에서 검은색으로 표현될 수가 있는데 이를 이용하여 조건 2를 적용하여 정상적인 2차 얼굴 영역의 원쪽, 오른쪽 지점을 얻어낼 수 있다.

조건 1. 원 폽xED일의 수가 1차 MBR 높이의 a이상이 되는 지점 (0<a<1).

조건 2. 조건 1을 만족하는 지점에서 중앙점까지, 검은 폽xED일의 수가 1차 MBR 높이의 b이상이 되는 지점이 있는 경우, 그 지점에서부터 조건 1을 만족하는 지점을 다시 검사 (0<b<1).
2차 얼굴 영역의 상단부 역시 조건 1을 변형하여 세로축으로 적용시키면 구할 수 있다. 이런 조건을 적용하여 얻은 2차 얼굴 영역은 머리카락 영역이 거의 제거되어 있음을 알 수 있고, 이후 비교적 쉬운 방법으로 눈의 위치를 찾아낼 수 있다.

(a) 검은색 머리 영상  (b) 금발의 얼굴 영상

그림 5. 머리카락 색이 서로 다른 영상의 1차 얼굴 영역 이차화 영상

2차 얼굴 영역은 머리카락이 거의 제거된 얼굴 영역을 포함한다. 얼굴 영역에서 눈과 눈썹은 다른 얼굴부분에 비해 상대적으로 아주므로, 2차 얼굴 영역의 영상을 가지고 다음과 같은 과정을 통하여 눈의 위치를 찾는다.

1. 2차 얼굴 영역을 이차화 시킨다.
2. 이차화 영상으로 y축 히스토그램을 구한다.
3. y축 히스토그램은 몇 개의 최대 점을 가지게 되는데 이들 점들 근처에 눈 위치의 근사치를 갖는 최대 점을 눈의 위치로 선택한다.

히스토그램을 이용하여 눈의 위치를 찾을 때, 눈의 위치는 2차 얼굴 영역의 좌측단 부에서 일정거리 내에 존재한다고 가정하며, 이마가 넓은 사람의 경우는 일정거리의 2층으로 조정한다. 이런 가정하에 눈의 위치를 찾아내면, 검출한 눈의 위치를 토대로 2차에 위치하여 아래쪽으로 일정치 이하가 되는 지점을 선택한다. 이 지점은 바로 눈의 아래쪽 경계선이 되는데, 이 지점을 선택하는 이유는 눈의 중앙점 혹은 위쪽 경계선은 눈의 폭과 길이가 크게 또는 작은 경우 그 위치가 상대적으로 아래쪽 경계선보다 많이 변하기 때문에 아래쪽 경계선을 선택했다.

(a) 2차 얼굴 영역  (b) 히스토그램 영상

(c) 눈의 위치 표시  (d) 정규화된 영상

그림 6. 눈의 위치 검출 영상

이런 방법을 이용하여 한 사람당 6장의 영상이 있는 120명의 영상, 720장에 대해서 실험한 결과 97.9%의 정확성 결과를 보이고 있다. 눈의 위치를 찾는데 실패한 영상은 두꺼운 겉안경 안경을 쓰고 있는 영상이 가장 많았고, 금발의 앞모 양상을이나 얼굴색이 전체적으로 두투은 영상도 다수 포함되었다.

3.3 면합수 모델을 이용한 얼굴영상의 특정추출

다양한 조명의 영향으로 같은 객체들이라도 다르게 보일 수 있다. 이러한 경우를 처리하기 위해서 기존에는 조명변화에 향상하게 하기 위한 전처리 기법을 도입하였다. 그러나 이러한 방법은 추가적인 계산비용이 소요되고, 조명에 따라서는 효과가 없거나, 왜곡 현상을 보이기도 한다. 이러한 문제들을 해결하기 위하여 얼굴영상으로부터 3차원적인 구조를 포함하는 특성치를 추출하여 돌글인식의 특성으로 사용한다.

얼굴 인식 시스템을 구비하기 위해서 본 논문에서는 원형 적합에 의한 방법을 사용하였다. 얼굴의 표현은 4면함수를 적합시켜 구한 구체의 방향값인 sina, cosa와 구체의 크기(magnitude)를 사용하며, 얼굴 영상들의 학습 집합을 이용하여 원형 얼굴을 구성한다. 원형 얼굴의 구조는 일반적으로 각 요소별 평균 값을 이용하지만, 가중치를 부여하는 방식을 적용할 수도 있다. 새로운 얼굴 영상이 입력되면, 데이터베이스에 저장되어 있는 원형 얼굴들과의 거리의 계산하여 가장 작은 거리를 가지는 원형 얼굴을 입력영상의 얼굴로 선택한다. 다차원 벡터 사이의 거리를 계산하는 방법은 여러 가지가 있지만, 여기서는 유클리디안 거리를 이용한다.

얼굴 표현으로 사용한 세 가지 요소 중 sina, cosa 값은 [-1, 1] 사이의 값이 가지는 반면, magnitude 값은 0-5 의 범위로 0에서 1000 정도의 값을 가지게 된다. 이 값들을 이용하여 유클리디안 거리를 구하게 되면, sina와 cosa의 두 거리 값에 비해 magnitude 값이 너무 크기 때문에 sina, cosa와
거리 값이 이동에서 변동되지 못한다. 따라서 magnitude에 의한 거리 값이 sinus와 cos 모의 거리 값과 비슷한 정도의 중요도를 갖게 하기 위해서 magnitude 값을 0 - 2 사이의 값으로 근사화 시켰다. 실험결과 5X5 크기의 경우 학습영상 전체에서 가장 큰 magnitude 값은 131이었으며, 원도우 크기의 크기 가 커짐에 따라 최대 magnitude는 작은 값을 보였다. 0 - 2 사이의 값으로 근사화 시키기 위해서 원도우 크기에 따라 magnitude의 최대값을 100, 80, 60, 40, 20이라고 가정하고 50, 40, 30, 10으로 나누었다. 이런 가정을 이용하면 모든 데이터가 0 - 2 사이의 값으로 바뀌지는 않지만, 실험결과 원도우 크기 5X5인 경우 실험 최대치인 150을 magnitude의 최대치로 가정하고 근사화 시킨 경우보다 100을 magnitude의 최대치로 가정하고 근사화 시켰을 때 더 좋은 인식 성능을 보였다.

[그림 7. 얼굴 특성 추출 과정]

본 논문에서는 두 얼굴 영상 사이의 거리를 유っくり 다른 거리의 평균값으로 표현하였고, 이를 D(f, f')라고 정의한다. 테스트 영상 f가 입력되면 데이터베이스에 있는 모든 얼굴 원형 D(f, f)를 구한다. 이 들 중에서 최소가 되는 D(f, f)를 구하고 이를 Dist는 다음과 같다.

$$\text{Dist} = \min_{f \in \mathcal{F}} D(f, f')$$

얼굴인식 과정에서 평균화에 대해서 정상적인 결과인지 아닌지에 대한 판단을 내려야 한다. 이때 오수락율(FAR: False Acceptance Rate)과 오거부율(FRR: False Rejection Rate)에 대한 연구가 필요하다. 오수락(FAR)이라면 다른 사람을 같은 사람으로 잘못 인식한 인식결과를 정상적인 인식 결과로 처리함으로 발생하는 오류이며, 오거부(FRR)로는 얼굴 인식 결과는 정상적인 결과이지만 검증결과에서 높은 오차값으로 인해 인식을 거부하는 오류이다. 이 두 가지 오류는 한쪽의 비율이 증가하면 다른 쪽의 비율은 감소하는 반비례의 성질을 가지고 있는데, 일반적으로 오인식 오류가 오거부 오류보다 심각한 장애를 일으키게 된다. 예를 들어 생체보안에 얼굴인식시스템이 사용되는 경우 오인식률이 높은 경우에는 인증이 허락되지 않아야 하는 외부인이 인증된 사람으로 오인되는 심각한 사례가 발생할 수 있다. 따라서 오거부율이 증가하더라도 오인식률을 낮추는 것이 더 중요하다.

인식결과에 대한 검증을 위해서 본 논문에서는 DFM (Distance Measure Factor)를 다음과 같이 정의하였다.

$$DFM = \frac{\text{Dist}}{\text{Dist}_f}$$

D는 원형얼굴들의 D(f, f)들을 중에서 두 번째로 작은 값을 의미하며, DFM은 인식 결과에 대해서 맞는지 틀린지를 판단하는 요소로 사용된다.

일반적으로 원형 경합은 특정 치를 추출하면 간단한 비교를 통해서 인식을 수행할 수 있고 데이터베이스에 새로운 원형얼굴을 구성하는 것이므로 새로운 클래식을 추가할 수 있다는 장점이 있는 반면, 인식하고자 하는 사람의 수가 늘어남에 따라 수행시간이 단점으로 작용한다는 단점이 있다.

3.4 변환수 특정 치를 입력으로 한 얼굴인식

얼굴인식 시스템을 구현하기 위해 먼저 원형 경합에 의한 방법을 구현해 보았다. 얼굴영상소의 학습 집합(training set)으로 학습을 이용하여 구체의 sin, cos가 구해지면, 이를 이용하여 원형 얼굴을 구현할 수 있다. 일반적으로 평균을 내어 구성하되, 중요한 특정 치를 선별하여 가중치를 더함 수도 있다. 이렇게 원형 얼굴을 구현한 후, 새로운 설정 영상이 입력되면, 변환수를 이용한 특정 치들 사이의 데이터베이스에 저장된 원형 얼굴들의 거리를 계산한다. 다차원 벡터사이의 거리를 계산하는 방법은 여러 가지가 있지만, 널리 쓰이는 유율리디안 거리를 이용한다. 그 식은 다음과 같다.

$$d^2 = (x - x')(x - x')$$

이렇게 각각의 원형 얼굴들과 비교하여 거리가 가장 가까운 얼굴을 인식된 결과로 출력하게 된다.

원형 경합 기법의 단점은 인식하고자 하는 얼굴의 수가 늘어날수록 수행시간이 허당으로 증가하게 되며, 가중치 조정이 어렵다는 것이다.

본 논문에서는 이러한 단점을 극복하기 위해, 원형 경합 이외에 신경망을 이용한 방법을 시도해 보았다. 신경망(backpropagation)을 사용하여, 얼굴영상에 대한 구체의 sin, cos 값을 입력으로 하고, 각 얼굴 당 하나의 출력노드를 구성한다. 신경망을 구성하는 방법은 여러 가지가 있다. 특히,
출액노드를 구성하는 방법에 대한 고찰이 필요하다. 출액노드를 한사람 당 하나의 출액노드로 구성할 수 있고, 출액노드를 이진화하여 구성할 수도 있다.

신경망의 연결선수가 늘어나면서, 적절한 학습에 필요한 폰트의 수도 늘어나므로, 출액노드의 수를 줄여 신경망을 단순화하기 위해서, 이전처럼 출액노드 구성이 더 낮지만, 한사람 당 하나의 출액노드로 구성하는 것이 실험 결과 더 좋은 인식율을 보였다. 다른 방법으로, 한사람 당 출액노드를 하나씩 구성하는 것이 아니라, 신경망을 하나씩 합당하는 방법도 제시된 바 있다[13].

그러나, 이 방법은 인식해야할 사람의 수가 늘어나면 따라 신경망 학습량의 구성을 시도해야 하므로, 상당한 비용이 드는 것으로 볼리지 않는 실험에선 제외하였다.

신경망의 입력은 -1, 1 사이의 값으로, 시그모이드 함수를 양극(bipolar)으로 결정하면, 입력에 넓아져 조정이 필요 없다. 학습용 영상에 의한 신경망의 학습이 완료되면, 새로운 실험 영상에 대해 가장 큰 출액노드를 인식할 결과를 출력하면 된다. 그림 8는 얼굴영상의 변환율을 이용한 특정치를 입력하여 다중 패스토론에서 인식시키는 과정이다.

그림 8. 다중 패스토론을 이용한 얼굴인식

4. 실험결과

본 논문에서 실험은 펌티엄 III-800Mhz, Windows 2000에서 Visual C++ 6.0을 사용하여 구현하였다. AR 얼굴 영상 데이터 베이스에서 한 사람 당 6장씩 120명의 얼굴영상, 720장의 검정을 이용하였고, 눈의 위치를 찾아 넘어 영상을 눈의 위치를 기준으로 좌측 70, 아래쪽 110, 폭 240의 영상(240X180)으로 잘라내고, 이를 80X60 크기의 영상으로 축소하여 얼굴 인식 실험에 사용하였다. 표 1은 성별, 눈 추출 성공률이다.

<table>
<thead>
<tr>
<th>구분</th>
<th>남자</th>
<th>여자</th>
</tr>
</thead>
<tbody>
<tr>
<td>눈 추출 성공률</td>
<td>98.2 (389/396)</td>
<td>97.5 (316/324)</td>
</tr>
</tbody>
</table>

표 1. 성별 눈 추출 성공률

얼굴인식에는 남자 61명, 여자 53명 총 114명의 얼굴 영상이 사용되었으며, 한 사람 당 3장은 학습용 영상으로, 다른 3장은 실험용 영상으로 이용하였다. 실험용 영상은 그림 9와 같이 표정 변화가 있는 영상, 왼쪽, 오른쪽 조명이 있는 영상 각 한 장씩 모두 3장을 사용하였다.

그림 9. 학습 영상과 실험 영상

모든 영상에 모폴로지 연산을 적용하여 잡음을 제거하고 원두구 크기에 입식율을 구하여 가장 적합한 크기를 얻었다. 표 2는 원두구별, 실험 영상별 인식 성능을 나타낸 것이다. 표정변화가 있는 영상, 왼쪽 조명이 있는 영상, 오른쪽 조명이 있는 영상, 세 가지 영상 전체에 대한 인식률이며, 5x5 크기의 원두구에서 가장 좋은 성능을 보인다.

<table>
<thead>
<tr>
<th>원두구 크기</th>
<th>얼굴표정 변화</th>
<th>왼쪽 조명</th>
<th>오른쪽 조명</th>
<th>전체</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 x 5</td>
<td>94.7</td>
<td>95.6</td>
<td>99.1</td>
<td>96.5</td>
</tr>
<tr>
<td>7 x 7</td>
<td>86.0</td>
<td>95.5</td>
<td>93.9</td>
<td>90.9</td>
</tr>
<tr>
<td>9 x 9</td>
<td>82.5</td>
<td>88.6</td>
<td>93.0</td>
<td>88.0</td>
</tr>
<tr>
<td>11 x 11</td>
<td>81.6</td>
<td>92.1</td>
<td>91.2</td>
<td>88.3</td>
</tr>
<tr>
<td>13 x 13</td>
<td>79.8</td>
<td>86.0</td>
<td>85.1</td>
<td>88.6</td>
</tr>
</tbody>
</table>

그림 10은 가변부분을 제외하지 않은 영상에서 PCA 방법과 sina, cosa 방법을 이용한 면함수 방법을 비교한 것이다. 면함수를 이용한 방법이 PCA 방법보다 성능과 인식시간에서 좋은 결과를 보였다.
그림 10. 가변 부분을 제외하지 않은 영상에서 얼굴 인식 성능 비교

그림 11은 가변 부분을 제외한 영상의 두 가지 면
합수 방법을 적용한 결과이다. 방법 1은 sina, cosa,
magnitude를 사용한 방법이고, 방법 2는 sina, cosa
만을 사용한 방법이다. sina, cosa를 사용한 면합수
방법은 PCA보다 높은 성능을 보였으며, sina, cosa만
을 사용한 면합수 방법보다 sina, cosa, magnitude를
사용한 면합수 방법이 더 좋은 성능을 보였다.

그림 11. 서로 다른 특성 틀을 이용한 면합수 방법
비교

그림 12는 가장 좋은 인식성능을 보인 원도우 크
기 5x5 인식 결과에 사용한 값을 X축으로 표준
DF로 표한 분포도이다. 그림 12에서
1은 인식에 성공한 영상들의 분포이며 2는 인식에
실패한 영상들의 분포이다. 인식에 실패한 영상들은
DF값이 거의 1에 가까음을 알 수가 있다.

그림 12. DMF와 Distj에 의한 분포도

표 3은 DMF값의 크기에 따른 오수라벨(FAR)과 오
거부율(FRR)을 나타낸 것이다. DMF값을 크게 설정
하면 오수라벨은 낮아지고 오거부율은 높아지며, 반
대로 DMF값을 작게 설정하면 오수라벨은 높아지고
오거부율은 낮아진다. 표3에서 보면 DMF가 1.06일
경우 실험 영상에 대해 오수라벨을 0.62이고, 오거부
율 2.6이며 이상적으로 인식 될 확률은 93.9이다.
DMF의 값을 적절한 수준으로 조정하여 원하는 정도
의 신뢰성과 인식성능을 얻을 수 있다. 본 실험에서
는 DMF값이 1.06일 때 오수라벨과 오거부율이 가
장 적절한 것으로 나타났다.

<table>
<thead>
<tr>
<th>DMF</th>
<th>설계 인식률</th>
<th>FAR</th>
<th>FRR</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00</td>
<td>96.5</td>
<td>3.5(12)</td>
<td>0.0(0)</td>
</tr>
<tr>
<td>1.02</td>
<td>95.3</td>
<td>2.3(8)</td>
<td>1.2(4)</td>
</tr>
<tr>
<td>1.04</td>
<td>94.4</td>
<td>1.8(6)</td>
<td>2.0(7)</td>
</tr>
<tr>
<td>1.06</td>
<td>93.9</td>
<td>0.6(2)</td>
<td>2.6(9)</td>
</tr>
<tr>
<td>1.08</td>
<td>92.1</td>
<td>0.3(1)</td>
<td>4.4(15)</td>
</tr>
</tbody>
</table>

표 3. DMF값에 따른 오인식률과 오거부율

5. 결론

본 논문에서는 눈의 위치를 찾아내어 이마로부터 틱
까지만 얼굴 영역을 추출하는 방법과 Haralick의 면
합수를 이용한 얼굴인식 방법을 함께 사용하였다. 이
들 통해서 얼굴이 아닌 부분을 얼굴 인식 과정에서
제외시켜 실험의 재현성을 높였다.

눈의 위치를 검출하기 위해서 먼저 1차 얼굴 영역
과 2차 얼굴 영역을 얻어내었다. 1차 얼굴 영역은 경
계선 정보를 이용하여 구하였고, 2차 얼굴 영역은 1
차 얼굴 영역을 이차로 시진 후, 검은 영역과 흰 영
역의 비율을 이용하여 구하였다. 이렇게 구한 2차 얼
Hee Sung Kim, Jong Ho Kim

금융 영역으로 Y축 허스토그램을 구하고 허스토그램의 최대값 정보를 이용하여 눈의 위치를 찾아내었다.

영구 인식 과정에서는 sina, cosα, magnitude 값으로 구성된 특정값만 추출해 내어 영구 인식의 입력으로 사용하였다. 그 결과 발행성의 비교가 용이하였고 입력노드를 위한 변환과정 없이 신경망의 입력으로 사용할 수 있었다.

제안된 방법을 형판정합의 방법과 신경망에서 인식하고 결과를 PCA방법과 함께 비교, 분석하였다. 변환수 특징치를 이용한 방법은 조정 보정을 위한 특별한 전처리 없이도, 조명변화에 대하여 96% 이상의 높은 인식률을 보였고, 이를 신경망에 적용한 결과 0.23초까지의 시간비용 절감의 효과를 보였다.

영구 인식 과정을 수행한 후 그 결과가 옳은지 그 문제에 대한 판단이 필요하다면 이를 위해서 DMF(Distance Measure Factor)를 정의하여 인식 과정의 결과를 검증하였으며, DMF값을 이용한 검증 과정에서 오인율과 오기에율을 구하여 가장 적절한 수준의 DMF값을 제시하였다.

눈의 위치를 찾아내는 과정에서 두꺼운 검은 대 안경과 금발의 머리카락에 의해 잘못된 위치를 찾는 경우가 있는데, 검은 대 안경과 금발의 머리카락도 눈의 위치를 정확히 찾아내는 방법이 연구되어야 한다. 본 논문에서는 각 관계별로 실험을 실시하였으나 실험영상에 대해서 각 관계를 합하여 적용하는 방법도 인식 성능을 향상시킬 수 있을 것이다.

변환수를 이용한 특징추출 방법은 얼굴 인식 이외에도, 조명에 진한 영상 인식이 필요한 다른 분야에서도 적용할 수 있을 것이다.

6. 참고 문헌


[12]. 김희승, 1993, 영상인식, 생능출판사, 서울.